DEFENSE ACQUISITION UNIVERSITY

F-38C AIRFRAME TESTING

CASE STUDY
PUBLISHED BY THE
DEFENSE ACQUISITION UNIVERSITY PRESS
FORT BELVOIR, VIRGINIA 22060-5565

DISCLAIMER

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328

F-38C AIRFRAME TESTING

Case Study

January 2003

Author(s) prepared this case as the basis for class discussion rather than to illustrate either effective or ineffective handling of a program management situation.

F-38C AIRFRAME TESTING

On 17 May 1997, Dennis Woltek, program manager (PM) for the F-38C fighter aircraft, was facing a tough decision on the ground test program. Woltek's chief engineer, Bob Horton, in collaboration with the F-38C contractor, had proposed deleting planned design load and fatigue testing of the tail section during ground testing. This proposal would save $4-6 million and reduce the program's critical path by three months. However, Ken Brown, his test director, believed the proposed cut in ground testing would expose the program to unacceptable risks. If Woltek wanted to proceed with the test he needed to make a decision by the end of the month to reserve the test site and ensure there was sufficient time for the contractor to build the tail assemblies.

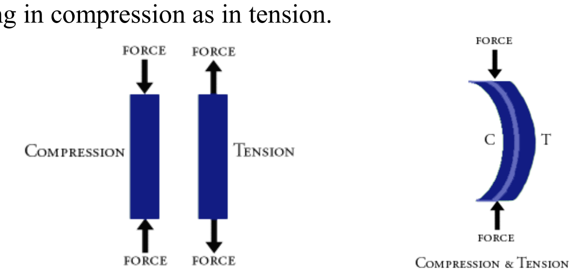
Test Planning

The F-38C was to be a "multi-role" version of the F-38. The "C" model was to be equipped with a larger engine and stronger wings to enable stores (ordnance, pods, etc.) carrying capacity as well as improved avionics to support ground attack capabilities. Because multi-role fighter operations would impose greater loads, other parts of the airframe besides the wings had to be structurally redesigned. The plan was to acquire 500 F-38Cs at a total acquisition cost of approximately $20 billion ($1B for development and $19B for procurement).

The developmental test program for the F-38C required a number of tests during a "ground" test phase before assembling prototypes for use during a "flight" test phase. The purpose of ground testing was to uncover design problems economically and safely prior to investing a substantial amount of funds and time assembling flight prototypes. This risk reduction was to be accomplished by testing various components and assemblies of the plane, including the airframe, to ensure they met design requirements and performed in a manner consistent with the modeling and simulation (M&S) tools used by the contractor in designing the "C" model.

The tail assembly on the "C" model was to be identical to those used on earlier versions of the F-38. Because of changes to other sections of the airframe and the addition of the ground attack missions to the plane's operational envelope, the load forces[1] exerted on the tail section of

[1] Forces come in two varieties: tension and compression. Tension is when you are pulling something apart, and thus stretching it longer. Compression is when you are pushing in on it, and thus are making it shorter. Drawn as an engineering force diagram, tension and compression look like the figures below. Materials used in constructing the F-38 airframe are equally strong in compression as in tension.



the "C" model would be different from those exerted on earlier models. The contractor's design engineers, however, had studied the M&S of the changes in load forces and did not consider the changes to be a problem. The M&S results clearly indicated no significant change in load force magnitude. Load forces had simply changed from being in tension to being in compression.

Brown had a great deal of confidence in the F-38 M&S tools; they had been validated and perfected to the "nth degree" over the years. He was concerned, however, that subtle details of the tail section omitted from the finite element analysis model of the F-38, such as the presence of screws or bolts, could result in actual load forces being significantly different from predicted load forces. In the instant case, Brown was uneasy with the model's assumption that a given level of load force on certain parts of the tail section would have the same impact on stress and fatigue whether exerted in tension or compression. Brown was worried that the presence of seemingly minor design details omitted from the model could cause a part to fail faster when in tension rather than compression and vice versa. For this reason, Brown believed there was a small but not insignificant probability that the expected changes in the load forces exerted on the F-38C tail section could result in unexpected structural failures or fatigue cracks during the life of the plane.

The consequence of a structural failure or fatigue crack in the tail section was potentially catastrophic. It could result in the loss of planes and pilots. The cost of correcting such a defect post-production, through rework, modification, or replacement, would likely range in the hundreds of millions of dollars, and/or the operational envelope of the plane could be severely restricted.

In preparing the developmental test plan in 1994-95, Brown had advocated reducing the risk of stress-related defects in the tail section by subjecting non-flying tail sections to design-limit load and fatigue testing representative of the F-38C mission profile.[2] This was to be done as part of ground testing. Horton had disagreed. In his mind the probability of stress problems in the tail section was too low to warrant the expense and added schedule time involved in conducting load and fatigue testing. The tail section had been tested extensively during ground and flight testing of earlier models of the F-38, and there had been no tail section stress problems experienced with the more than 600 F-38s that had been put in service.[3] Horton had been more concerned about focusing test and evaluation (T&E) resources on areas of the plane that were being changed.

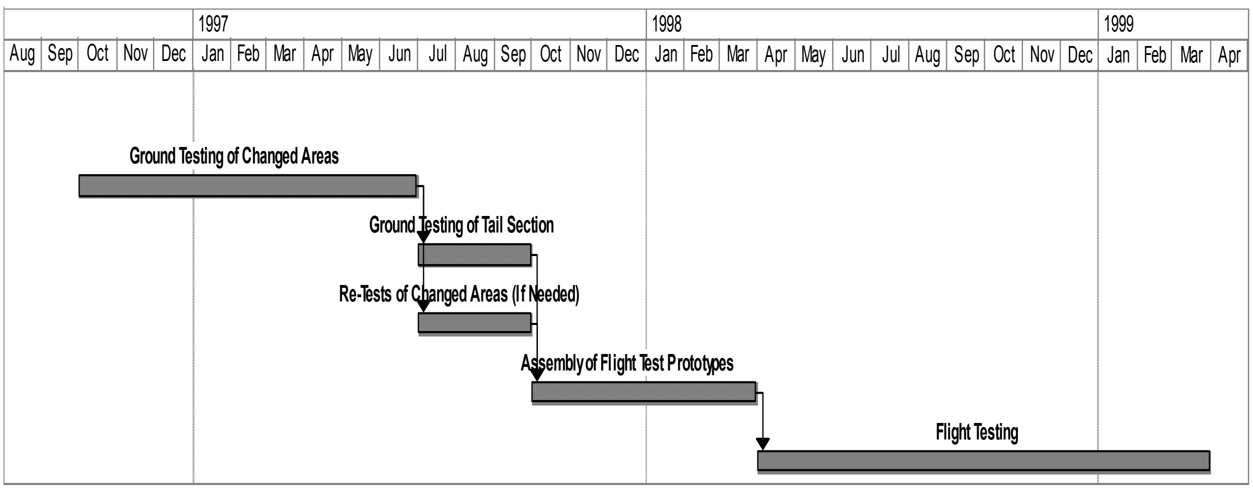
[2] To determine if an airframe part meets load force requirements, the part is subjected to design-limit load testing. In design-limit load testing, a non-flying airframe part (wing, tail section, etc.) is instrumented and subjected to certain tests, such as bending by mechanical means, to determine the part's structural strength. Typically parts must demonstrate 150 to 200 percent of the design-limit load (the highest estimated load the aircraft is expected to ever experience in flight) during static testing depending on the type of aircraft. To determine service life, airframe parts are subjected to fatigue testing. In fatigue testing, a non-flying part is instrumented and subjected to repeated loads by mechanical means, simulating various flight conditions and payloads. Parts must typically survive testing equal to four lifetimes to demonstrate a service life, i.e. a part must survive 120,000 hours of fatigue testing to demonstrate a 30,000-hour service life.

[3] In ground testing of earlier F-38 models the tail section passed a 32,000 hour fatigue test and withstood forces up to 220% of the design load limit before failing.

During test planning, Woltek had not weighed in on the tail section testing issue specifically, but he was adamant that ground testing not exceed the 12 months planned in the integrated program schedule. A longer ground test phase would drive up program costs, delay initial operational capability (IOC), and possibly require significant reprogramming of funds.

After consulting with the T&E and risk management integrated product teams (IPTs), Brown and Horton had arrived at a solution that both found acceptable. They agreed that load testing of the tail assembly would be scheduled as the last critical path event for the ground test phase. Testing of those areas of the plane that were undergoing design changes (e.g. engine, wings, and avionics) would be given priority on testing resources since both believed there was a greater probability of design problems with the changed areas than with the tail section. Corrections of any design problems with changed areas could then be worked concurrent with tail section testing, which was expected to yield peace of mind for Brown and no design issues for Horton.

F-38C Test Schedule



Test Execution

Ground testing had started October 1996 and was scheduled to be completed by the end of September 1997. For various reasons—delays in building assemblies, delays in getting access to test facilities, unexpected technical glitches, etc.—ground testing was more than three months behind schedule and $10 million dollars over budget (contractor cost related to schedule delay and direct testing costs). Most of the problems occurred during the first months of testing. After January 1997 things improved, but by May it was clear that the ground phase would not be completed until the end of the calendar year.

In response to these developments, Woltek had requested that his staff and the contractor review ground and flight test plans and determine if there were ways to make up the three-month slip and reduce testing costs without exposing the program to unreasonable risks. Woltek did not want the program to be showing significant cost and schedule problems this early in the testing program if possible. He knew such problems would serve as an invitation for additional outside help in running his program.

Horton's Proposal

In early May 1997 Horton had informed Brown that he intended to propose to Woltek that the Government delete design load and fatigue testing of the tail section during ground testing to save time and money. Horton and the contractor's engineers believed that in light of current budget and schedule constraints, extensive ground testing of the tail assembly—a low risk area—was no longer affordable. They argued that although stress data collected during upcoming flight testing would not be as robust as data collected during ground testing, the data would nevertheless be adequate to confirm the integrity of the design. Brown disagreed with this proposal. The test budget and schedule had already been cut several times during the three years the F-38C program had been in existence. From Brown's perspective, any further cuts in testing would expose the program to unacceptable risks. His preference was for Woltek to seek additional funding or cut other areas of the program's schedule and budget.

After lengthy discussion, Brown and Horton concluded that they were both in general agreement concerning most of the facts related to Horton's proposal but had different judgments concerning the reliability of the finite element analysis model and what was an acceptable risk. They agreed to jointly present their positions on the proposal to Woltek.

Woltek had been favorably impressed with the presentation. Brown and Horton had both advanced good arguments in support of their positions. They had presented the following information:

- Deleting the tail-section ground test would shorten the ground phase by three months and allow building of the flight test prototypes to begin in October as originally planned. It would also reduce test-related costs by $4-6 million: $1 million from avoiding the direct cost of conducting the test (test site expenses and cost of test articles) and $3-5 million in reduced contractor costs derived from shortening the length of ground testing from 15 to 12 months and the overall program schedule by three months.
- Horton believed the probability that the M&S had not uncovered a major tail section stress problem (i.e., had indicated a false pass) was on the order of 1-2%. Brown on the other hand, believed there was a 5-10% probability that the M&S had not uncovered a major tail section stress problem. Both acknowledged that their probabilities were subjective estimates based on professional judgment and experience working with modeling and simulation. Woltek had great confidence in the professional integrity and technical competency of both men and accepted their judgments as sincere representations of their points of view.
- Brown and Horton judged that there was a 1-5% probability of failing to detect a major tail section stress problem not already indicated by M&S during flight-testing. Both agreed that collection of stress data during flight-testing was of minimal cost ($100 thousand) and not a schedule driver since the data would be

collected incidental to the performance of other flight test events. Both also agreed that flight testing would be less robust than ground testing in detecting stress problems.

- In their judgment there was a 0.1-0.5% probability of not detecting a major tail section stress problem not already indicated by M&S during ground testing. Brown and Horton both believed that testing the tail section during the ground phase would provide extremely robust design data, far superior to what could ever be ascertained from data collected during flight testing. During ground testing, the tail section would be exercised to the breaking point, thus revealing the design's load and fatigue limits. Stress data collected during flight testing could be used to tweak and validate M&S and possibly reveal stress problems within portions of the operational envelope flown during testing, but would not yield definitive information on design load and fatigue limits.
- In their judgment the probability of not detecting a major tail section stress problem (not already indicated by M&S) if both test events (ground and flight) were conducted was effectively zero (0.001-0.025%).
- If major tail section stress problems were detected during flight testing it would cost about $25-50 million to fix and re-test the prototypes. The test program would also incur a schedule delay of 3-6 months.
- If major tail section stress problems were detected after the production and fielding of 500 planes it would cost about $500 million-$1 billion to fix the planes, and/or the warfighter would have to make do with a less capable (i.e., lower g's and reduced stores carrying capacity) and less durable plane.

		Risk of Defect If No Ground Test		Expected Value of Outcome	
DEFECT DETECTED DURING	**OUTCOME**	**Horton**	**Brown**	**Horton**	**Brown**
Flight Testing	$25-50M 3-6 Mo Delay	1-2%	5-10%	$250K - $1M	$1.25M - 5M
Post Production (After Flight Testing)	$500M-$1B Flight Restrictions	0.01 - 0.1%	0.05 - 0.5%	$50K - $1M	$250K - $5M

Horton had advanced six basic reasons for deleting ground testing of the tail assemblies. First, circumstances, namely cost and schedule constraints, had changed significantly since test planning was conducted in 1994-95. As a result of current cost and schedule over-runs, Horton argued that it was necessary to ask whether the program could afford ground testing the tail section simply to reduce a risk from very low to nil. Skipping ground testing and relying solely on flight testing to reveal any stress problems not detected by M&S would reduce the "worst case" risk of an undetected problem to a range of 0.1% (Horton's worst case) and 0.5% (Brown's worst case).[4] That put worst case odds between 1 in 200 and 1 in 1,000.

Second, when the test plans and integrated program schedule were originally put together, testing of the low risk tail assembly was scheduled for the end of the ground phase. It was contemplated that any fixes and re-testing necessary to resolve design issues identified during testing of the higher-risk "changed areas" would be worked concurrent with tail assembly testing. It was for this reason that Horton had eventually acquiesced on Brown's desire to ground test the tail assembly before building flight test prototypes. As things had turned out thus far, there did not appear to be any design issues that absolutely had to be resolved prior to commencing with assembly of the test prototypes. Therefore, elimination of tail section testing would allow the program to complete ground phase testing and commence with assembly of flight test prototypes by the end of September as originally planned.

Third, on the basis of expected value it was difficult to justify ground testing the tail section prior to assembling the flight prototypes. Horton, using his probabilities for the existence of a defect not already indicated by M&S, had calculated the expected value (assuming no ground testing) of the cost to fix prototypes if a stress problem was detected during flight-testing as being in the range of $250K to $1M.[5] It was Horton's position that the program would be better off self-insuring against the small probability of incurring $25-50 million in prototype re-work expense and a 3-6 month delay rather than paying $4-6 million for the ground tests and adding three months to the schedule simply to reduce the probability of incurring the loss to effectively zero.[6] Horton had argued that gambles of this nature were implicit in many F-38C program decisions and although one or two might turn out unfavorably, overall the program benefited in terms of cost, schedule, and performance from bearing risks on favorable terms. He also stated that ground testing of the tail section could not be justified on the basis of expected value even if Woltek was inclined to accept Brown's judgment of a 5-10% probability that M&S had failed to indicate a defect in the tail section. Even under Brown's worst case probability (10% chance of M&S not indicating a defect) and worst case outcome ($50 million) would the decision to do or not do ground testing become an even gamble.

Fourth, Horton contended that stress data collected during flight testing would sufficiently reduce the risk of an undetected defect in the tail section to a level that would make deletion of ground testing a favorable gamble on the basis of expected value. Using his probabilities of an undetected defect after completion of flight testing, Horton calculated the expected value of the cost to fix the F-38C fleet as being in the range of $50K to $1M.[7] Again, Horton argued that only under Brown's worst case probabilities and outcomes would the decision whether to do ground testing become an even gamble.

[4] Horton's worse case of undetected defect (false pass) after M&S and flight testing: .02 x .05 = .001. Brown's worse case: .1 x .05 = .005

[5] .01 x $25 million = $250K. .02 x $50 million = $1 million.

[6] At worst, a 0.5% chance of ground testing not detecting a defect for which there was a 2% chance that it was not detected by M&S. In other words, the odds would be 1 in 100,000 (.005 x .02 = .00001).

[7] Lower Value: .01(M&S) x .01(Flight) x $500 Million = $50K; Upper Value: .02(M&S) x .05 (Flight) x $1 Billion = $1 Million.

Fifth, the F-38C test and evaluation master plan (TEMP) did not mention ground testing of the tail section. Nor did the F-38C developmental test and evaluation (DT&E) plan approved by the Office of the Secretary of Defense's DT&E oversight office, the Deputy Director for Strategic and Tactical Systems, Developmental Test and Evaluation (DDST&S (DT&E)), specifically require ground testing of the tail section. The F-38C DT&E plan explicitly called for ground testing of the wings and other areas of the air frame undergoing change. Ground testing of unchanged areas was to be conducted "as needed" as determined by the PM. To Horton, this indicated that the DT&E community did not view ground testing of the tail section as necessarily being consistent with the best use of the program's T&E resources. The fact that ground testing the tail section was beneficial was merely a necessary condition for doing the test and not a sufficient condition. Spending time and money to test the tail section would mean less time and money would be available to satisfy other testing needs of the program. It was Horton's opinion that the benefits of ground testing the tail section was relatively low and the resources could be put to better use addressing unknown unknowns that would no doubt surface during developmental testing.

Sixth, Horton suggested that perhaps Brown was somewhat biased towards testing things that the T&E community knew how to test well, such as the structural integrity of air frames, while paying less attention to things that it did not know how to test as well, such as avionics software. Horton had expressed concern that the program was falling into the trap of testing for reasons of orthodoxy rather than risk management. He averred that some in the test community seemed to have a preternatural disposition towards eliminating virtually all risks associated with performance requirements testers knew how to test while accepting significant risks with performance requirements for which well developed testing methodologies did not exist.

Brown had offered three basic reasons for why ground testing should be conducted. First, Brown believed it was unwise to invest the time and money building flight prototypes when there was a significant probability of the existence of tail section design problems undiscovered through M&S. It was better to invest $4-6 million and three months of critical path time confirming the integrity of design at this stage than taking the risk that a problem would crop up during flight testing. Even worse was the possibility of that the tail section would pass flight testing with an undetected defect that wouldn't show up until the fielded planes had accumulated significant service life.

Second, although Brown acknowledged the utility of expected value analysis he argued that other factors such as program reputation and risks to pilots had to be considered. He questioned whether the program could survive if tail section problems surfaced during flight testing resulting in $25-50 million in additional cost and 3-6 months of delay. Furthermore, he did not believe expected value analysis was very useful when dealing with catastrophic outcomes the program could not self-insure itself against, i.e., $500M to $1B to fix 500 planes with defective tail sections.

Third, Brown stated that it was prudent to eliminate almost all risk of airframe defects. For the F-38C to have a long service life it was necessary to make sure the airframe was rock solid. As Brown had explained, the life of the F-38C could always be extended with upgrades to the engine, avionics, etc.; but such upgrades would not be feasible unless the airframe was highly durable. It was extremely difficult and almost always uneconomical to modify or rebuild an airframe. Brown offered the long life of the B-52 as an exemplar of the benefits of a durable airframe.

Woltek observed that during the presentation Horton had emphasized the probability of the F-38C not passing ground loads testing was very low and thus cast doubt on the utility of investing money and time in this testing. Brown on the other hand, did not dispute Horton's confidence in the F-38C's tail section passing the flight testing but seemed more concerned that the F-38C could pass flight testing but still have a potentially catastrophic design flaw. To Brown, ground phase testing was necessary to reduce the probability of a false pass to an acceptable level.

Decision

Woltek did not enjoy having to make decisions of this type. He understood all too well, however, that making such decisions was an important aspect of his PM duties. This would be just one of the many risk management decisions he had made thus far on the program and he knew there would be many more.

As Woltek ruminated over the decision, the specter of the C-5A program loomed in his mind. Woltek did not want that type of disaster on his hands. To meet an aggressive late-1960s delivery schedule, the C-5A program was executed using McNamara-era acquisition initiatives such as "total package procurement" and "concurrent development and production." One consequence of this approach was that design-limit load and fatigue testing of the C-5A did not occur until production was well underway—it was impossible to make test articles and complete design limit load and fatigue testing without delaying production milestones by months. Hence, when it was discovered that the wings failed at 126% of design limit load and cracks appeared during fatigue testing after less than half the 30,000-hour service life, there was little that could be done but impose severe payload restrictions on an airlifter that was being procured specifically for enhanced range and payload capability. The wings of the airlifter were eventually modified during the 1970s at cost of $1.5B, allowing the airlifter to meet payload performance requirements.

Woltek realized that the circumstances of F-38C program were vastly different than the C-5A. The C-5A program was a full scale development effort of an entirely new plane in an era before the advent of powerful M&S design tools such as finite element analysis. In contrast, the F-38C was a variation of an existing aircraft for which there existed an abundance of test and performance data. The risks and unknown unknowns associated with the structural integrity of the F-38 airframe were immensely smaller than those related to the C-5A during its development. Nevertheless, the saga of the C-5A did serve as a stark reminder to Woltek that structural problems do in fact occur.

Woltek could not dismiss Brown's concerns about the possibility of structural problems that had gone undetected by M&S. He recalled that the C-17 airlifter had been designed in the late 1980s and early 1990s with all the benefits of contemporary M&S design tools. The M&S results clearly indicated that the C-17 wing design would meet the 150% design limit load. Yet, when tested in 1992, the wing failed at 119% of the design limit load. The wing design had to be improved and retested before the program proceeded to production. Again, however, the C-17 had been a completely new plane and not a variation of an existing plane.

Woltek understood there was never enough time and money to test everything. He knew the world would little note a decision not to ground test the tail section if things turned out well. On the other hand, if there were bad consequences such a decision would no doubt become part of the lore of the acquisition community and frequently cited as an example of imprudent program management. He could just imagine some acquisition academic droning on about how a reckless PM screwed up the F-38C program by trying to take shortcuts.

Although there were no areas of the F-38C program schedule or budget that could easily serve as a bill payer to offset the program’s current cost and schedule overruns—all the low hanging fruit had been plucked by his predecessors—Woltek was confident that he could forge an alternative solution (take a little here, cut a little there, compress some here, etc.) if ground testing the tail section was truly the right thing to do. Of course the program would have to pay the price of such a decision by bearing the risks associated with ratcheting up the tautness of the program’s budget and schedule. Although the tail section test decision was not overly momentous in light the overall size of the program, Woltek understood that he could tinker with the budget and schedule for only so much before putting program execution in serious jeopardy. More importantly, he knew there were unknown unknowns that would appear in the future presenting additional budget and schedule challenges.

Woltek wished he knew how the costs and risks associated with the tail section decision compared with those of the yet unknown future budget and schedule decisions he would have to make. Then he could better decide whether to do the test or assume the risk of not doing the test and preserve “tinkering” space for future contingencies. Although experience and history afforded him a vague notion of the range of possibilities it was impossible to have such foreknowledge. To Woltek it was not just the risks associated with the tail section but also the irreducible uncertainty of the future that made his decision so vexing. He fretted over the possibility of finding himself confronting future cost and schedule overruns and regretting he had not been more aggressive in saving money and time by assuming risks such as those presented by the tail section. Or would he find himself contending with performance problems late in the program that could have been resolved at a far lower cost if they had been identified early in test.

After further deliberation, he decided he would….

F-38C AIRFRAME TESTING

Student Assignment

1. Drawing from your own knowledge and experience can you think of other pros and cons to Horton's proposal not mentioned in the case?
2. Does Woltek have any other alternatives than the two presented by Brown (do the design load and fatigue testing on the tail section prior to building the flight test prototypes) and Horton (delete the tail section ground test from the test program entirely)?
3. If you were in Woltek's position what criteria would you use for evaluating the alternatives?
4. If you were in Woltek's position what decision would you make?

5. Prepare an "ISSUE SHEET" (one page)
   a. Select an important issue in the case.
   b. Provide a short background summary. (Summarize the issue and briefly explain why it is important.)
   c. State your recommended solution for resolving the issue.